Accurate clinical genetic testing for autoinflammatory diseases using the next-generation sequencing platform MiSeq

AUTHOR(S):

Nakayama, Manabu; Oda, Hirotsugu; Nakagawa, Kenji; Yasumi, Takahiro; Kawai, Tomoki; Izawa, Kazushi; Nishikomori, Ryuta; Heike, Toshio; Ohara, Osamu

CITATION:

Nakayama, Manabu ...[et al]. Accurate clinical genetic testing for autoinflammatory diseases using the next-generation sequencing platform MiSeq. Biochemistry and Biophysics Reports 2017, 9: 146-152

ISSUE DATE:

2017-03

URL:

http://hdl.handle.net/2433/227669

Biochemistry and Biophysics Reports 9 (2017) 146–152

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

Accurate clinical genetic testing for autoinflammatory diseases using the next-generation sequencing platform MiSeq

Manabu Nakayama[a,b,*], Hirotsugu Oda[c], Kenji Nakagawa[c], Takahiro Yasumi[c], Tomoki Kawai[c], Kazushi Izawa[c], Ryuta Nishikomori[c,*], Toshio Heike[c], Osamu Ohara[a,d]

[a] Department of Technology Development, Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
[b] Laboratory of Pharmacogenomics, Graduate School of Pharmaceutical Sciences, Chiba University, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
[c] Department of Pediatrics, Kyoto University Graduate School of Medicine, 54 Shogoin Sakyo, Kyoto 606-8597, Japan
[d] Laboratory for Integrative Genomics, RIKEN Center for Integrative Medical Sciences (IMS), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

ARTICLE INFO

Keywords:
Next-generation sequencing (NGS)
Somatic mosaicism
Primary immunodeficiency diseases (PIDs)
Amplicon sequencing
Multiplex PCR

ABSTRACT

Autoinflammatory diseases occupy one of a group of primary immunodeficiency diseases that are generally thought to be caused by mutation of genes responsible for innate immunity, rather than by acquired immunity. Mutations related to autoinflammatory diseases occur in 12 genes. For example, low-level somatic mosaic NLRP3 mutations underlie chronic infantile neurologic, cutaneous, articular syndrome (CINCA), also known as neonatal-onset multisystem inflammatory disease (NOMID). In current clinical practice, clinical genetic testing plays an important role in providing patients with quick, definite diagnoses. To increase the availability of such testing, low-cost high-throughput gene-analysis systems are required, ones that not only have the sensitivity to detect even low-level somatic mosaic mutations, but also can operate simply in a clinical setting. To this end, we developed a simple method that employs two-step tailed PCR and an NGS system, MiSeq platform, to detect mutations in all coding exons of the 12 genes responsible for autoinflammatory diseases. Using this amplicon sequencing system, we amplified a total of 234 amplicons derived from the 12 genes with multiplex PCR. This was done simultaneously and in one test tube. Each sample was distinguished by an index sequence of second PCR primers following PCR amplification. With our procedure and tips for reducing PCR amplification bias, we were able to analyze 12 genes from 25 clinical samples in one MiSeq run. Moreover, with the certified primers designed by our short program—which detects and avoids common SNPs in gene-specific PCR primers—we used this system for routine genetic testing. Our optimized procedure uses a simple protocol, which can easily be followed by virtually any office medical staff. Because of the small PCR amplification bias, we can analyze simultaneously several clinical DNA samples with low cost and can obtain sufficient read numbers to detect a low level of somatic mosaic mutations.

1. Introduction

There are many similar diseases that present clinically with fever and inflammation, and many of these occur due to gene mutations. Traditionally, identifying disease-causing gene mutations has been time-consuming and costly, requiring patient samples to be analyzed by a specialized laboratory employing sophisticated equipment. In recent years, however, a push is being made to make genetic testing available in clinical practice, in settings wherein doctors can quickly determine whether the patients in front of them have a mutation of genes known to underlie certain diseases. This type of clinical genetic testing plays an indispensable role in providing patients with quick, definite diagnoses. In order to reduce the economic burden on patients, clinical genetic testing must be performed accurately and quickly at a relatively low cost.

A general trend in Next Generation Sequencing (NGS) platforms is a move from the Roche GS 454 (its manufacturer no longer offers technical support) to the Illumina HiSeq. Development of a simple, accurate, and high-throughput method is embodied by the MiSeq platform. Presently, many kits are commercially available, but they are still too expensive for routine sequencing. A significant contributor to the cost of genetic testing is labor costs. Most currently used sequen-

* Corresponding authors.
E-mail addresses: nmanabu@kazusa.or.jp (M. Nakayama), rnishiko@kuhp.kyoto-u.ac.jp (R. Nishikomori).

http://dx.doi.org/10.1016/j.bbrep.2016.12.002
Received 14 July 2016; Received in revised form 22 November 2016; Accepted 9 December 2016
Available online 23 December 2016

cers require highly trained technicians to analyze patient samples and operate sequencers. As labor costs occupy a large percentage of the total cost of genetic testing, it cannot be completely ignored when considering the overall cost of genetic testing. Introducing a simple testing protocol would be most effective in reducing labor costs, because it would eliminate complicated and extensive training necessary for a technician to become sufficiently skilled to prepare, analyze, and run samples. This would be especially helpful for genetic diseases that physicians are beginning to see more frequently in the clinic.

Autoinflammatory diseases are a relatively new category of disease proposed in 1999 and are thought to be caused by mutations of genes involved in the innate immune system, but generally not in the adaptive immune system. Autoinflammatory diseases often are characterized by recurrent fever and inflammation. Almeida de Jesus and Goldbach-Mansky have categorized these into six groups based on clinical manifestations [1]. Familial Mediterranean Fever (FMF); TNF receptor-associated periodic syndrome (TRAPS); mevalonate kinase deficiency (MVK)/hyperimmunoglobulinemia D with periodic fever syndrome (HIDS); cryopyrin-associated periodic syndromes (CAPS); neonatal-onset multisystem inflammatory disease (NOMID)/chronic infantile, neurological, cutaneous; and articular (CINCA) syndrome are all examples of autoinflammatory diseases. Somatic mosaicism of NLRP3 in patients with CINCA have been reported [2–5].

Clinical genetic testing should be performed carefully because autosomal dominant gain of function mutations are reported in many of these autoinflammatory diseases. Moreover, it is not known exactly whether somatic mutations other than those in NLRP3 confer one of the autoinflammatory diseases.

We have chosen two-step tailed PCR amplification and amplicon sequencing for library preparation for the MiSeq platform for five reasons. Firstly, around 10 genes have been generally estimated as being candidate genes responsible for autoinflammatory diseases based on symptoms associated with these diseases. Secondly, new target genes will likely be added as this field of study continues to progress. Even if one made custom-ordered probes or capturing probes for all target genes simultaneously, researchers will want additional probes for newly discovered responsible genes in the future. It is costly to custom order a new set of probes again, because commercially available systems generally do not support small additions and changes to existing probes. Thirdly, high sensitivity and sufficient read depth are necessary in order to detect somatic mosaic mutations. Fourthly, as simple a procedure as possible is desirable. Lastly, costs need to be reduced. Motivated by these five reasons, we attempted to simultaneously analyze several clinical DNA samples on one MiSeq run, aiming to decrease PCR amplification bias as much as possible.

2. Materials and methods

2.1. Patients and clinical diagnosis

A total of 108 patients having an autoinflammatory disease diagnosis were consecutively diagnosed and recruited at the Department of Pediatrics, Kyoto University Graduate School of Medicine. All of the patients were Japanese and provided written informed consent (below) for inclusion in the high-throughput sequencing analysis.

2.2. Ethics statement

All patients provided written informed consent after we gave them a full explanation of the study. All patients gave us explicit permission to analyze their DNA sequencing data for genes responsible for autoinflammatory diseases. This study was approved by both the Human Research Ethics Committee of the University of Kyoto and the Kazusa DNA Research Institute.

2.3. DNA samples

DNA samples were de-identified with regard to subjects' personal information. DNA from patients' blood was purified using a QIAamp DNA blood kit (QIAGEN, Venlo, Netherlands). Before use, DNA samples were quantified by Qubit Fluorometric quantitation (Thermo Fisher Scientific, Waltham, MA, USA).

2.4. Design PCR primers

Primers were basically designed by ExonPrimer Perl script (https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html). The design input parameters were 30 bp, 150 bp, and 20 bp for the minimal distance between the primer and exon/intron boundary, maxima target size, and overlap, respectively. The standard values were selected according to the parameters of Primer3Web options (http://bioinfo.ut.ee/primer3/) [6].

ExonPrimer suggested candidate paired PCR primers. We wrote a short custom program to detect common SNPs in the DNA sequence of candidate PCR primers. This program consulted the UCSC In Silico PCR database (http://rohsdb.cmb.usc.edu/GBshape/cgi-bin/hgPcr) and the dbSNPs database (dbSNP 135). We named this short program "Primer-dbSNP search" (i.e., SNPs finder for PCR primers). After checking common SNPs using Primer-dbSNP search, we redesigned undesirable primers to avoid common SNPs using commercially available oligo 6.0 Primer Analysis Software (Molecular Biology Insights, Inc., Colorado Springs, CO, USA). M13FW (5′-TGTAAAACGACGGCC −3′) and M13RV (5′- GGAAACAGCTATGAC −3′) were added at the 5′ end of each forward and reverse primer, respectively. Primers were synthesized by Eurofins Genomics K.K. (Tokyo, Japan).

2.5. Library preparation following multiplex PCR

Supplemental Table 1 lists 234 PCR primer pairs used to target the protein coding region and a minimum of 20 bp of untranslated, flanking intronic region of the genes. Dual-indexed secondary PCR primers are described in Fig. 1. For the first PCR amplification, amplification was performed in a 0.2 ml 8-strip PCR tube (Corning, Corning, NY, USA) in a final volume of 50 μl. It had 25 μl of 2 X Multiplex PCR Buffer (Mg^{2+}, dNTP plus) (TAKARA, Multiplex PCR Assay Kit Ver. 2); 0.25 μl of Multiplex PCR Enzyme Mix; each gene-specific primer pool (0.05 μM); and 400 ng of human genomic DNA. The following amplification steps were performed: 94 °C for 1 min, 10 cycles at 94 °C for 30 s, and 60 °C for 1 min, followed by incubation at 72 °C for 10 min. The resulting PCR products were purified twice with AMPureXP beads (Beckman Coulter Inc., Brea, CA USA). To minimize accidental cross-contamination between samples, we purified the PCR product using a low-binding tube, not a 96-well plate. Typically, 10 ng of purified PCR product was used for each PCR step. This is equivalent to using only one-third of the eluate from AMPureXP purification. For secondary PCR amplification, amplification was performed in a final volume of 50 μl. It had 25 μl of 2 X Multiplex PCR Buffer (Mg^{2+}, dNTP plus) (TAKARA, Multiplex PCR Assay Kit Ver. 2); 0.25 μl of Multiplex PCR Enzyme Mix; each index primer (10 μM; each D501-D508-like and D701-D712-like Dual-indexed secondary PCR primers); and 10 ng of purified PCR product. The following amplification steps were performed: 94 °C for 1 min, 5 cycle at 94 °C for 30 s, 55 °C for 10 s, and 72 °C for 30 s, followed by incubation at 72 °C for 10 min. The PCR product was purified twice with AMPureXP beads using non-skirted thin-wall 96-well 0.2 ml plates. The purified PCR product was quantified with the Kapa Library Quantification Kit for the Illumina NGS (Kapa Biosystems, Wilmington, MA, USA), using an ABI 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific). After quantification, we typically mixed equal molar concentrations of PCR product from 25 individual DNA samples, applied it to MiSeq using the MiSeq Reagent

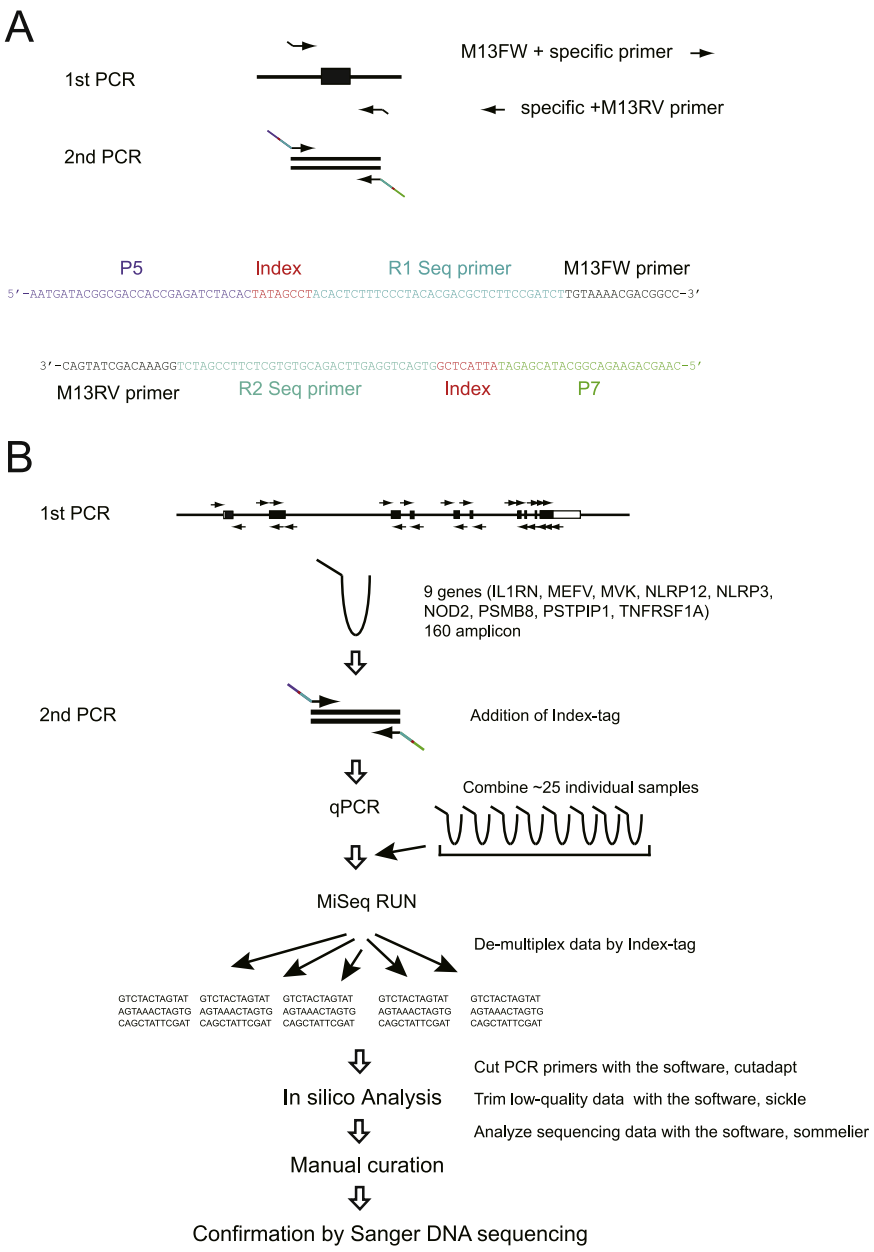

Fig. 1. Schematic diagram illustrating our MiSeq approach to identifying genes responsible for autoinflammatory diseases. (A) First gene-specific primer and second primer for two-step tailed PCR. For the first PCR amplification step, the first PCR primer was composed of M13FW 15mer and a gene-specific primer, and M13RV 15mer and gene-specific primer. For the second amplification step, the second PCR primer (forward orientation) was composed of P5, an index tag, an R1 seq primer, and an M13FW DNA sequence. The second PCR primer (reverse orientation) was composed of P7, an index tag, an R2 Seq primer, and an M13RV DNA sequence. (B) Scheme of clinical genetic testing using the MiSeq platform. All coding exons from the nine genes analyzed in this study were amplified simultaneously by multiplex PCR in one tube. After the second PCR amplification step, index tags were added to distinguish 25 individual samples. After the MiSeq run, the resulting data were demultiplexed using the index tags. Sequencing data were analyzed in silico. The software program Sommelier detects variants and creates a file annotating identified variants. After manual curation, important mutations, e.g., missense mutations, were confirmed by Sanger DNA sequencing.

kit v3 600 cycles (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. In line with the so-called DarkCycle procedure, we did not collect data for the first 15 cycles, because the first 15 bp are common sequences containing M13FW and M13RV DNA sequences in R1 and R2 reads, respectively.

2.6. Data analysis

Each PCR primer sequence at the 5′ terminus of the read sequence was trimmed from the sequence read data, and each PCR primer and additional M13FW or M13RV 15-mer sequence in the 3′ terminus of the read sequence were also trimmed from the sequence read data by the software Cutadapt [7]. Low quality sequences, namely low quality 25 (Q25) from the 3′ terminus of the read sequence were trimmed from the sequence read data without the primer sequence using the software sickle [8]. The software's quality filter removes the R1 and R2 reads, which contain low quality sequences at a high ratio. The software Sommelier, described previously [5], is a variant caller program that includes Blat software [9] as a component. Trimmed sequence read data were mapped onto DNA sequence data of the amplicon derived from the human reference genome (hg19) using Blat and default parameters. Variants, which were identified by Sommelier, were annotated using a short program (proc_exome_mutation_b37.v4.pl). After manual curation, all missense, nonsense, and frame-shift mutations, as well as other severe mutations, were confirmed by Sanger sequencing using BigDye Terminator v3.1 (Thermo Fisher Scientific)

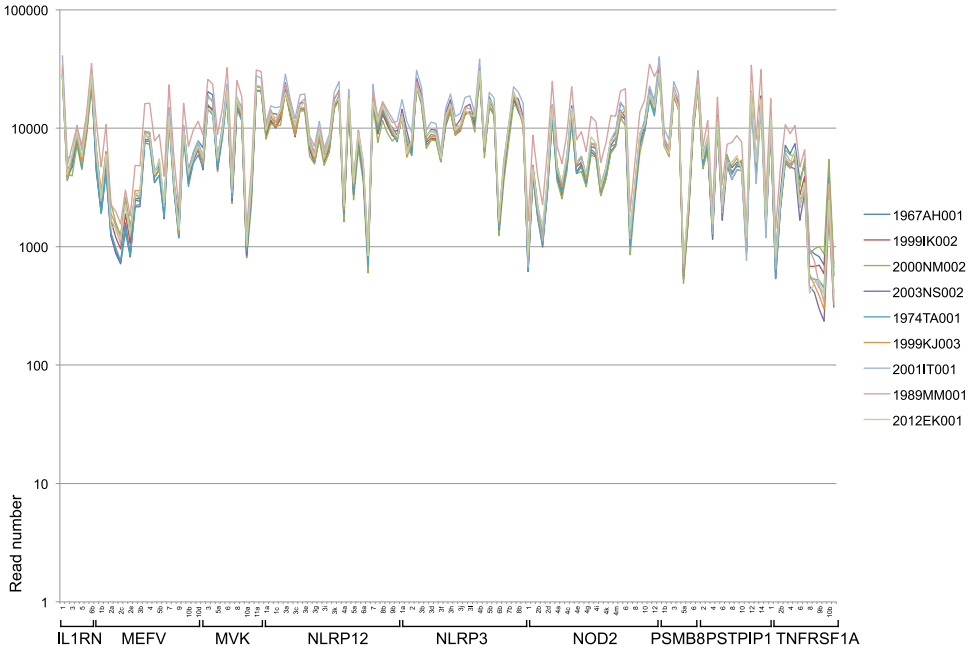

Fig. 2. Read number comparisons for each amplicon derived from multiplex PCR amplification. Most amplicons were read over 1000 times. The average read number was 9052.

and ABI 3130 or 3710 instruments (Thermo Fisher Scientific).

3. Results and discussion

As illustrated in Fig. 1A, we chose two-step tailed PCR for high-throughput analysis using the Next Generation Sequencing platform, MiSeq, to detect variants in several clinical DNA samples. Because the read number obtained in one MiSeq run is limited and cannot be increased easily, we sought to minimize bias of PCR amplification and deviation of read number of each amplicon by analyzing several DNA samples simultaneously.

First, we selected nine genes (IL1RN, MEFV, MVK, NLRP12, NLRP3, NOD2, PSMB8, PSTPIP1, and TNFRSF1A) associated with autoinflammatory diseases, and made a total of 160 primer pairs. These nine genes contain 88 coding exons. Since large exons need multiple primer pairs, we made 320 primers for PCR amplification. At present, the maximum read length of MiSeq is 300 bp (2×300) for paired-end sequencing, since our criteria for clinical genetic testing is that each base should be read in both directions. Large coding exons completely overlapped with some amplicons derived from split exons. These gene-specific PCR primers contained an additional common 15 bp of M13FW and M13RV sequence in the 5′ terminus to be amplified during secondary PCR amplification.

Multiplex PCR was performed in one tube using all gene-specific primers. Because PCR amplification bias increases during each step of the PCR cycle, in our protocol we performed 10 cycles for the first PCR amplification and 5 cycles for the second PCR amplification. Various PCR amplification parameters were assessed during preliminary experiments in order to determine the optimal conditions (i.e., the number of PCR cycles, the concentration of each PCR primer in the reaction, and the annealing temperature) for the gene-specific primer sets for the nine genes responsible for autoinflammatory diseases (see Methods). We noticed that performing the AMPureXP purification produced better results for MiSeq DNA sequencing, since small PCR products derived from primer dimers produce many empty sequence reads, which waste large volumes of limited and valuable read numbers in MiSeq.

Secondary PCR amplification was performed to add an index sequence distinguishing each PCR product derived from each clinical DNA sample. Typically, the indexed PCR product derived from the 25 individual DNA samples were mixed and applied to one MiSeq run. After a MiSeq run, sequencing data were demultiplexed by using the index sequence, PCR primer sequences and low quality regions were trimmed, and sequence reads containing low-quality sequences were removed. This was accomplished using the variant call software Sommelier, which calls variants from the trimmed sequence read data and writes a user-friendly file containing a variant list with annotation information in Excel format.

Supplemental File 2 shows typical clinical genetic testing results for autoinflammatory diseases. Sheet 1, “Read Count” of Supplementary File 2 shows the read number of each amplicon. These indicate that there are enough read numbers in each amplicon. The average read number for forward and reverse orientation is 4583 and 4469, respectively. The rate of over 20 reads for both orientations is 100%. The rate of over 30 and 100 reads in both orientations are 99.79% and 98.95%, respectively. Moreover, total read numbers, or the sum of forward and reverse read numbers for each amplicon, are shown in Fig. 2. Average total read number of each amplicon was 9052. The average minimum read number and maximum read number was 405 and 32434, respectively. The ratio of the average maximum read number to average minimum read number was 79.94. We observed that the total read number in some amplicons was relatively small, and they contained an unusually high GC-rich region. PCR amplification bias is unavoidable for such extremely GC-rich regions. In conclusion, in our multiplex experimental condition, PCR amplification bias is low, which permits parallel analysis of many clinical samples.

The read number for each amplicon in Fig. 2 and Supplemental File 2 relates to the average depth of each exon; that is, the quotient is total base number derived from all read numbers divided by base number of the exon. Because large exons were covered by sequence reads of some overlapping amplicons and gaps in exons, special attention was required, especially for large exons. The read depth for each base was completely checked for all coding exons of the targeted genes. When no gaps are present in the coding exons, MiSeq achieves sufficient read depth for each base. For regions sequenced at a depth greater than 25, the depth coverage was 100%; for those sequenced at a depth greater than 30, the depth coverage was 99.98%. These results showed that amplicon sequencing with sufficient depth was performed.

Sheet 2 of Supplementary File 2 (Final Results sheet) shows a list of mutations identified using the MiSeq platform. The list contains

genome position, gene symbol, mutation type, frequencies for forward orientation and for reverse orientation, coding DNA sequence (CDS) level change, protein level change, AA variation, dbSNPs, and allele frequency in dbSNPs for each mutation. Such variants lists that contain annotation information are very easy to use, especially when many genes and many clinical samples are analyzed simultaneously.

In the on-going scientific discussion about the sequence of PCR products, especially those concerning exome sequencing using the capture method, the view is sometimes espoused that artificial mutations are incorporated during PCR amplification. However, we believe this is not relevant in the case of our approach for the following reasons. First of all, we used 400 ng of human DNA, which is equivalent to 1.3×10^5 copies of the haplotype, which translates to an average read number of about 9000 per amplicon. This read depth is sufficient enough so that one can ignore any rare, artificial mutations that theoretically can occur during PCR amplification. Because our PCR products were derived from amplification of an independent template of 1.3×10^5 copies, even if misincorporation did occur during the first PCR amplification step, its effect would have been extremely small or negligible.

Frequency of variants is an important numerical aspect for determining whether a mutation is a homozygous or heterozygous mutation. In the early stages of this study, we noticed a case that had a variant frequency of less than 12% in one amplicon of a sample from one patient. Sanger DNA sequencing revealed that genomic DNA corresponding to the gene-specific PCR primer, by chance, contained a SNP. The presence of an accidental SNP in a gene-specific PCR primer causes a one-base mismatch between the PCR primer and the DNA template. In heterozygotes, the DNA template from one allele hybridizes with the primer completely. On the other hand, a DNA template with a SNP hybridizes with the mismatch. Therefore, PCR amplification efficiency is much lower in cases of mismatch than it is when the PCR primer and DNA template complete match. Fig. 3 shows the relationship between the location of a mismatch in a PCR primer and the detection frequency of a variant allele containing a SNP. As expected, a SNP near the 3′ end of the primer greatly affects the detection frequency of the variant. Especially having a SNP at the first and second base position in a PCR primer significantly decreases the detection frequency by 1–2%.

Although SNPs very rarely occur in gene-specific PCR primers, because we plan to analyze genes of several hundred patients, we took countermeasures to reasonably guard against this problem. We wrote a short program to detect SNPs in gene-specific PCR primers. When common SNPs (> 1% frequency) were detected in a PCR primer, the PCR primer was re-designed to avoid SNP sites. Completely new PCR sets void of common SNPs were used in subsequent MiSeq experiments.

Our genetic testing using MiSeq also detected large deletions in X-chromosome-linked genes from a patient (data not shown; manuscript in preparation) on other similar panel sets for primary immunodeficiency diseases, but not on the panel we used for the present study to examine autoinflammatory-disease-causing genes. The significant decrease in read number of the amplicon indicates the deletion of the exon(s). In fact, we did confirm that this was due to a large deletion in the patient's genome.

Next, we tried to detect mosaic mutations using our MiSeq detection system. In the first model experiments, two DNA samples with different SNPs were mixed at a rate of 5% and 10%, respectively. This mixed DNA sample was used for our MiSeq experiment. When a SNP in the material was a heterozygote, as expected we detected nearly 2.5% and 5% of variant frequency. This demonstrates that variant frequency values derived from our system are extremely precise. In the next experiment involving the detection of mosaic mutations, we tested three real DNA samples, which were identified to have somatic mosaic mutations. As shown in Table 1, our MiSeq experiment detected 35.3%, 7.0%, and 6.3% of the variant frequency of these samples.

Next, we tested whether we could add the primers of three other genes onto the nine-gene panel used for our MiSeq experiment. The gene-specific primers of three genes, NLRC4, PLCG2, and HMOX1, were designed according to the procedure described above. We took special precautions to avoid accidental incorporation of common SNPs into the designed candidate primers. PCR primer pairs of 74 amplicons derived from 45 coding exons in these three genes were added to the pool of PCR primers used for the previously identified nine genes responsible for autoinflammatory diseases. Addition of the three new primer pairs did not negatively affect the accuracy of genetic testing using our MiSeq system. Indeed, our system is sufficiently robust and flexible that it could accommodate changes, such as the addition of these three genes. One strong advantage this system has in terms of clinical genetic testing is its flexibility, which would be especially useful for relatively “new” syndromes like autoinflammatory diseases. These diseases have only recently been recognized, and as they receive more attention from the scientific community, likely more responsible genes will be identified and will need to be characterized.

The mutations that we found in this study are summarized in Table 2. Twenty-seven missense mutations emerged from the 108 patients examined. The most notable of these were Gly566Ala in NLRP3, Lys34Thr and Asp369Gly in PSTPIP1, Arg410His and Pro115Arg in MEFV, and Gln902Lys in NOD2, which have not been reported previously. All missense mutations identified in this study were confirmed by Sanger sequencing. Independent DNA sequencing using gold-standard methods, such as Sanger sequencing, validates the reliability of clinical genetic testing using the MiSeq system in identifying mutations. In addition, independent DNA sequencing would identify any errors that may have arisen from, for example, a technician inadvertently mixing up patient samples.

In some of the samples we analyzed, we did not detect any mutations in genes that were identified previously to cause autoinflammatory diseases [1]. This suggests that not all responsible genes were included in our PCR-primer sets, because some genes responsible for autoinflammatory diseases have not been identified yet. Our aim, therefore, is to broaden our ability to test for more autoinflammatory disease-associated genes simultaneously using our MiSeq platform.

In the present study, we noticed that MiSeq had some tendencies that could lead to inaccurate results if not fixed via in silico analysis. First, in the eight serial thymine (T) cluster, we often observed artificial T deletions, at an approximate frequency of 2%. Second, mis-synthesized PCR primers were often detected as noise in the analysis. The DNA synthesizer we used never produced 100% exact oligomers. Rather, it mis-synthesized a very small volume of product. Moreover, cutadapt software hardly ever detected and trimmed PCR primers containing a mutation. Thus, the noise derived from mis-synthesized primers should be removed during in silico analysis. Fourth, pseudo genes, gene families, and repeat sequences located extremely close to an exon can lead to the production of false amplicons during PCR, which, when amplified, produce noise reads. False mapping was detected as noise. Such false mapping should also be removed by in silico analysis.

In our MiSeq approach, the primer design step is one of the most important steps to ensure successful multiplex PCR amplification with minimal PCR amplification bias. PCR primers that overlap each other must be avoided, above and beyond the need to make good PCR primers.

Exome analysis can survey mutations in all exons without one knowing ahead of time the genes responsible for any given disease. At the present time, however, exome analysis remains expensive and has a relatively low read number. Indeed, with a small read depth, even the MiSeq platform cannot distinguish true mutations from misreads and surely cannot detect low frequencies of somatic mosaic mutations.

Our MiSeq approach is suitable for routine clinical genetic testing. If our MiSeq approach does not detect a significant mutation, if necessary, additional analysis (e.g., exome analysis) should be per-

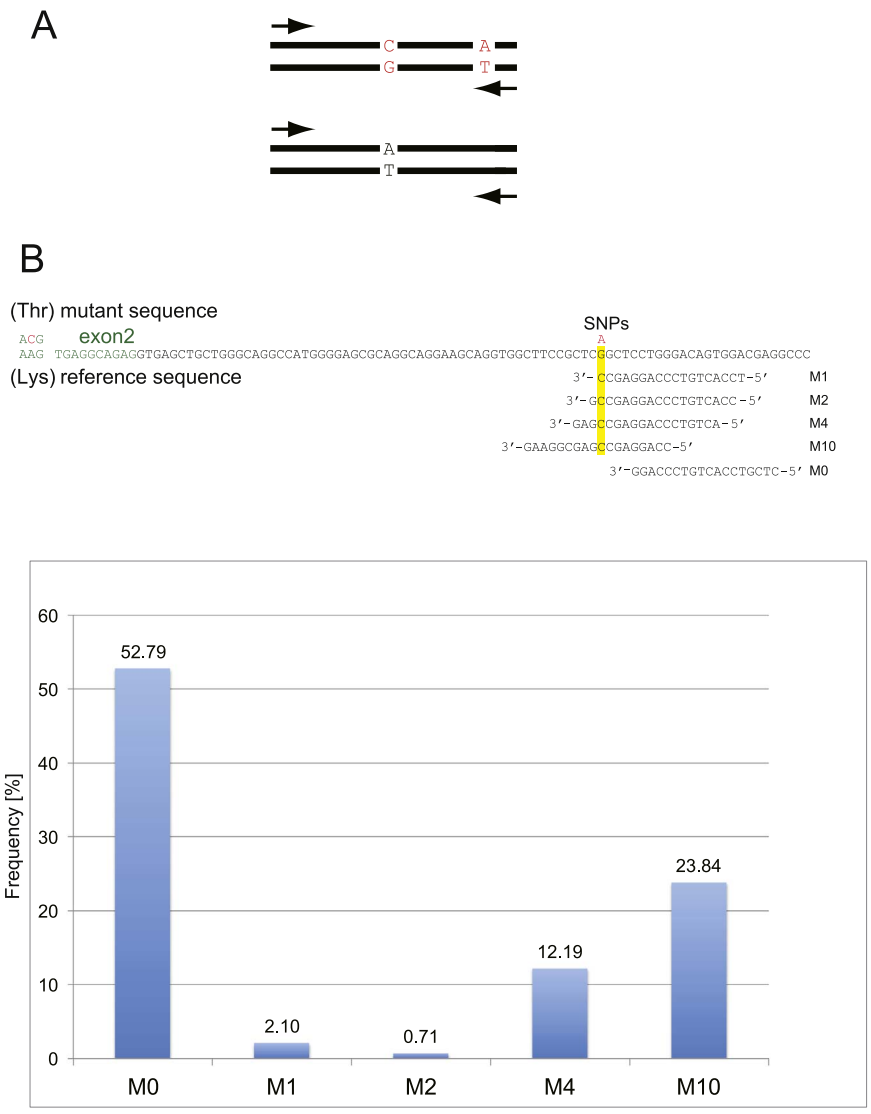

Fig. 3. Effect of amplicon read number in MiSeq on PCR amplification using genomic DNA containing SNPs. (A) The presence of a SNP in a PCR primer causes unbalanced PCR amplification of each allele, causing the products derived from those alleles to have different read numbers. The two black bars represent double-stranded DNA from one allele. One allele (bottom illustration) has the same sequence as the reference sequence (A type). The other allele (top illustration) contains a C variant in the middle and an A variant (SNP) in the primer. During the hybridization step, a mismatch occurs at the SNP between the PCR primer and the template DNA, obstructing further PCR amplification of the DNA containing the SNP. The resulting PCR products could lead one to under- or overestimate the frequency of mutations. (B) How location of SNPs in PCR primers can affect the detection frequency of mutations. One allele has two mutations in the middle of the exon in the area of the PCR primer site. M1 has one mismatch in the 3′-terminal end. M2, M4, M10 have the same mismatch but it is located in the second, fourth, and tenth positions, respectively, from the 3′-terminal end of the PCR primer. M0 does not have any mismatches in the area of the SNP in the gene-specific primer site. One allele has a C-to-A mutation, resulting in a Thr-to-Lys substitution in exon2. Gene-specific primer pools lacking M0, M1, M2, M4, M10 were used for multiplex PCR amplification after completing the MiSeq procedure, and then data were analyzed by a variant caller program, Sommelier. Mismatches occurring near the 3′ terminus of primers significantly decrease the frequency at which mutants are detected.

formed. We are currently developing another gene panel for the MiSeq platform in order to analyze another category of primary immunodeficiency disease. Along this line, we are developing a system to analyze simultaneously over one thousand amplicons of 57 genes that confer primary immunodeficiency diseases, although we need to reduce PCR amplification bias and differences in read number in NGS.

Table 1
Detection of somatic mosaic mutations in patient DNA using our MiSeq analysis system.

Patient ID	Genome Position	Gene Symbol	Fwd Freq	Rev Freq	Fwd Read	Rev Read	Region	CDS Level Change	AA Variation	dbSNP
11028IS	chr1:247587751	NLRP3*	35.767	34.922	4985	2643	exon3e	c.1000 A > G	p. Ile334Val	–
11040GM	chr1:247588450	NLRP3*	7.161	7.226	5837	4954	exon3i	c.1699 G > A	p. Glu567Lys	rs104895389
11040GM	chr1:247588450	NLRP3*	6.758	7.068	5993	3551	exon3j	c.1699 G > A	p. Glu567Lys	rs104895389
12039IA	chr1:247588450	NLRP3*	6.603	6.141	6542	5699	exon3i	c.1699 G > A	p. Glu567Lys	rs104895389
12039IA	chr1:247588450	NLRP3*	6.127	6.286	7002	4216	exon3j	c.1699 G > A	p. Glu567Lys	rs104895389

In "Regions" having large exons, the exons were overlapped by amplicons. Letter after exon number indicates the name of each amplicon. The names of the amplicons are identical to the names of the corresponding primers, as shown in Additional File 1. Amplicons, exon3i, and exon3j were independently amplified, sequenced, and analyzed.

* In NLRP3, the clinical genetics community typically uses the second Met as the initial start site, not first the Met. Thus, in this study, NLRP3 had two additional amino acids because the first Met was used for the annotation of the NLRP3 gene.

Table 2.
Variants identified in the genetic testing carried out in this study.

Genome Position*	Gene Symbol	CDS Level Change	Protein Level Change	AA Variation	dbSNP	Number of carriers
chr1:247582310	NLRP3**	c.214 G > A	Missense	p. Val72Met	rs117287351	2 heterozygotes
chr1:247586640	NLRP3**	c.392 A > G	Missense	p. Lys131Arg	rs188623199	1 heterozygote
chr1:247587659	NLRP3**	c.914 A > C	Missense	p. Asp305Ala	rs180177447	1 heterozygote
chr1:247588067	NLRP3**	c.1322 C > T	Missense	p. Ala441Val	rs121908146	1 heterozygote
chr1:247588442	NLRP3**	c.1697 G > C	Missense	p. Gly566Ala		1 heterozygote
chr1:247597508	NLRP3**	c.2431 G > A	Missense	p. Gly811Ser	rs141389711	1 heterozygote
chr12:110013879	MVK	c.155 G > A	Missense	p. Ser52Asn	rs7957619	1 heterozygote
chr12:6442956	TNFRSF1A	c.269 C > T	Missense	p. Thr90Ile	rs34751757	2 heterozygotes
chr15:77310553	PSTPIP1	c.101 A > C	Missense	p. Lys34Thr		1 heterozygote
chr15:77324670	PSTPIP1	c.773 G > C	Missense	p. Gly258Ala	rs34240327	2 heterozygotes
chr15:77328263	PSTPIP1	c.1106 A > G	Missense	p. Asp369Gly		1 heterozygote
chr16:3293405	MEFV	c.2082 G > A	Missense	p. Met694Ile	rs28940578	1 homozygote, 3heterozygotes
chr16:3297095	MEFV	c.1508 C > G	Missense	p. Ser503Cys	rs190705322	3 heterozygotes
chr16:3299462	MEFV	c.1229 G > A	Missense	p. Arg410His		1 heterozygote
chr16:3299468	MEFV	c.1223 G > A	Missense	p. Arg408Gln	rs11466024	9 heterozygotes
chr16:3299586	MEFV	c.1105 C > T	Missense	p. Pro369Ser	rs11466023	9 heterozygotes
chr16:3304158	MEFV	c.910 G > A	Missense	p. Gly304Arg	rs75977701	3 heterozygotes
chr16:3304463	MEFV	c.605 G > A	Missense	p. Arg202Gln	rs224222	5 heterozygote
chr16:3304626	MEFV	c.442 G > C	Missense	p. Glu148Gln	rs3743930	5 homozygotes, 23 heterozygotes
chr16:3304724	MEFV	c.344 C > G	Missense	p. Pro115Arg		2 heterozygotes
chr16:3304739	MEFV	c.329 T > C	Missense	p. Leu110Pro	rs11466018	12 heterozygotes
chr16:50753909	NOD2	c.2704 C > A	Missense	p. Gln902Lys		1 heterozygote
chr19:54301639	NLRP12	c.2785 G > A	Missense	p. Ala929Thr	rs146368839	2 heterozygotes
chr19:54304482	NLRP12	c.2755 C > T	Missense	p. Arg919Trp	rs61741349	1 heterozygote
chr19:54313707	NLRP12	c.1206 C > G	Missense	p. Phe402Leu	rs34971363	2 heterozygotes
chr19:54327313	NLRP12	c.116 G > T	Missense	p. Gly39Val	rs34436714	11 heterozygotes
chr6:32811629	PSMB8	c.145 C > A	Missense	p. Gln49Lys	rs2071543, rs147533146	16 heterozygotes

* Genome position was based on human hg19 as a reference.
** In NLRP3, the clinical genetics community typically uses the second Met as the initial start site, not the first Met. Thus, NLRP3 has two additional amino acids because the first Met was used for the annotation of the NLRP3 gene in this study.

4. Conclusions

In this study, our MiSeq approach using multiplex PCR can perform genetic testing with high sensitivity and accuracy to detect even low-level somatic mosaic mutations that cause autoinflammatory diseases. This simple procedure can be both time- and cost-effective, and can be conducted by technicians with an average skill level and without special training. We routinely analyze several clinical DNA samples in one MiSeq run every two weeks. This simple procedure will also minimize inevitable human errors and guarantee the accuracy required for diagnostic application.

Conflict of interest

The authors declare that there are no conflicts of interest.

Authors' contributions

M.N. carried out the MiSeq experiments; M.N., H.O., and O.O. carried out the bioinformatics analyses; K.N., T.Y., T.K., K.I., R.N., and T.H. contributed research resources; M.N., R.N., and O.O. conceived the study, and participated in its design and coordination. All authors helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank all patients who participated in this study. We are grateful to K. Sato, M. Takazawa, K. Sumi, and T. Watanabe at the Department of Technology Development, Kazusa DNA Research Institute for their technical assistance. This work was supported by grants from the Kazusa DNA Research Institute, and in part, by JSPS KAKENHI Grant Number 24310143 of the Grant-in-Aid for Scientific Research (B).

Appendix A. Transparency document

Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.bbrep.2016.12.002.

References

[1] A. Almeida de Jesus, R. Goldbach-Mansky, Monogenic autoinflammatory diseases: concept and clinical manifestations, Clin. Immunol. 147 (3) (2013) 155–174. http://dx.doi.org/10.1016/j.clim.2013.03.016.
[2] M. Saito, R. Nishikomori, N. Kambe, A. Fujisawa, H. Tanizaki, K. Takeichi, et al., Disease-associated CIAS1 mutations induce monocyte death, revealing low-level mosaicism in mutation-negative cryopyrin-associated periodic syndrome patients, Blood 111 (4) (2008) 2132–2141. http://dx.doi.org/10.1182/blood-2007-06-094201.
[3] R. Goldbach-Mansky, Current status of understanding the pathogenesis and management of patients with NOMID/CINCA, Curr. Rheumatol. Rep. 13 (2) (2011) 123–131. http://dx.doi.org/10.1007/s11926-011-0165-y.
[4] N. Tanaka, K. Izawa, M.K. Saito, M. Sakuma, K. Oshima, O. Ohara, et al., High incidence of NLRP3 somatic mosaicism in patients with chronic infantile neurologic, cutaneous, articular syndrome: results of an international multicenter collaborative study, Arthritis Rheum. 63 (11) (2011) 3625–3632. http://dx.doi.org/10.1002/art.30512.
[5] K. Izawa, A. Hijikata, N. Tanaka, T. Kawai, M.K. Saito, R. Goldbach-Mansky, et al., Detection of base substitution-type somatic mosaicism of the NLRP3 gene with > 99.9% statistical confidence by massively parallel sequencing, DNA Res.: Int. J. rapid Publ. Rep. Genes Genomes 19 (2) (2012) 143–152. http://dx.doi.org/10.1093/dnares/dsr047.
[6] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, et al., Primer3 – new capabilities and interfaces, Nucleic Acids Res. 40 (15) (2012) e115. http://dx.doi.org/10.1093/nar/gks596.
[7] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnetjournal 17 (1) (2011) 10–12.
[8] J. NA, F. JN, Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at⟨https://github.com/najoshi/sickle⟩, 2011.
[9] W.J. Kent, BLAT – the BLAST-like alignment tool, Genome Res. 12 (4) (2002) 656–664. http://dx.doi.org/10.1101/gr.229202 (Article published online before March 2002).